ANALYSIS OF HIV PROTEINS USING DSP TECHNIQUES

Irena Cosic

Bioelectronics Group, Department of Electrical & Computer Systems Engineering
Monash University, Clayton, 3168, Australia
Phone: +61 3 9905 5356 Fax: +61 3 9905 3454
e-mail: Irena.Cosic@eng.monash.edu.au

Abstract:-Our previous studies suggested that digital signal processing methods can be used to analyse linear sequences of amino acids to reveal the functional informational within the protein sequence. In this study both spectral and time-frequency methods are applied to the analysis of the functional content of HIV virus envelope proteins. Here, we have identified specific RRM frequency of HIV proteins, predicted active sites in these proteins and compared these predictions with experimentally determined active sites.

I INTRODUCTION

The Acquired Immuno-Deficiency Syndrome, AIDS, is caused by the human immunodeficiency virus (HIV) which predominantly attacks cells having CD4 molecule on their surface. The mechanisms of viral entry and replications are already well known but the mechanism by which cell depletion occurs is still poorly understood and thus there is still no efficient cure for the disease. The first step in the infection of host cells by HIV virus is interaction between HIV envelope protein and the CD4 antigen. HIV envelope protein, gp160 is initially divided into two fragments proteins gp120 and gp41, which then bind to each other to interact with the CD4 cell antigen. This, as any other protein functional interaction, is a specific and selective interaction. The rules governing the coding of the protein's ability to selectively interact with other molecules is yet to be discovered. The Resonant Recognition Model (RRM) [1-4] is one attempt to identify the selectivity of protein interactions within the amino acid sequence. The RRM proposes that the specificities of protein interactions are based on the resonant electromagnetic energy transfer at the specific frequency for each interaction. The RRM is applied here to analyse HIV virus proteins, their interactions and in particular HIV active sites within HIV envelope proteins.

Studies on antiviral activity of peptides from HIV-1 envelope revealed three peptides: DP-107, peptide-637-666 and the most potent, T20, all from gp41 have antiviral activity [5-7]. The mechanism by which these peptides inhibit HIV-1 infection is still not known. We have previously analysed inhibitory activity of these peptides, using the RRM model [1,2,15]. Here we use the continuous wavelet transform to predict positions of inhibitory peptides as well as to predict the position of binding site between HIV-env and CD4 antigen in the HIV envelope protein.

II METHODS: the Resonant Recognition Model (RRM)

The *Resonant Recognition Model* is based on the finding that there is a significant correlation between spectra of the numerical presentation of amino acid and their biological activity[1-4]. By assigning the *electron-ion interaction potential* (EIIP) value [8] to each amino acid, the protein sequence can be converted into a numerical sequence. These numerical series can then be analysed by appropriate digital signal processing methods (fast Fourier transform is generally used). To determine the common frequency components in the spectra for a group of proteins, the multiple cross-spectral function was used. Peaks in this function denote common frequency components for the sequences analysed. Through an extensive study, the RRM has reached a fundamental conclusion: **one RRM characteristic frequency characterises one particular biological function or interaction [1-4]**. Thus, it can be postulated that RRM frequencies characterise not only general function but also recognition and interaction between particular protein and its target

Once the RRM characteristic frequency for a particular biological function or interaction has been determined, it is possible to identify the individual amino acids so called "hot spots", or domains that contribute mostly to the characteristic frequency and thus to protein's biological function as well. Initially, these amino acids were identified using Inverse Fourier Transform (IFT) [1-4]. Recently, the wavelet transform (WT) has been incorporated into the RRM. The wavelet transform uses a set of dilated and translated wavelets as the signal decomposition basis [9]. The continuous wavelet transform(CWT) of signal *s(t)* is defined as:

cwt(a,b) = ò***s(t) (1/***Ö***a)*** *y* ***[(t-b)/a] dt***

where *b* is the shift factor and *a* is the scale factor. CWT provides same time/space resolution for each scale and thus, CWT can be chosen to localise individual events,

0-7803-7211-5/01$10.00©2001 IEEE

Report Documentation Page

Report Date 25 Oct 2001	**Report Type** N/A	**Dates Covered (from... to)** -

Title and Subtitle Analysis of HIV Proteins Using DSP Techniques	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Bioelectronics Group Department of Electrical & Computer System Engineering Monash University Clayton, 3168 Australia	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor’s Acronym(s)**
	Sponsor/Monitor’s Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Oct 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 4	

such as the active site identification. The active sites along the protein sequence are determined through studying the set of local extrema of the moduli in the wavelets transform domain. Sites with energy concentrated local extrema are the locations of sharp variation points of EIIP and are proposed to be the most important information sites[10,11,16].

III PRELIMINARY RESEARCH

The RRM model is based on the finding that distribution of delocalised electron energies along the protein play crucial role in determining protein biological activity [1-4]. In fact, it was found that proteins having the same biological function (same target or receptor) share same periodicities (frequencies) in this energy distribution. These frequencies are denoted as characteristic frequencies for the particular biological process. Although receptors share the same characteristic frequency with ligand proteins, indicating that their recognition is on the basis of periodic matching between energy distribution, the phase at particular frequency is opposite (phase shift close to $\underline{+}\pi = \underline{+}3.14$ rad between receptors and ligands). [1-4]

Envelope gp160 proteins from 19 different HIV isolates (lav1/bru, hxb2, nl43, sf2, sc, mn, rf, wmj2, cdc451, ny5, jh3, brva, eli, mal, z6, z2z6, z3, z321, jy1, ndk, oyi) were analysed by the RRM procedure. Only one common frequency component for all analysed isolates was obtained as prominent peak at frequency $f1=0.185\underline{+}0.001$ (signal-to-noise=484) in the crosspectral function (Fig.1). According to RRM concepts, it can be proposed that this frequency characterises the common biological behavior of all analysed proteins[1-4,12,13,15].

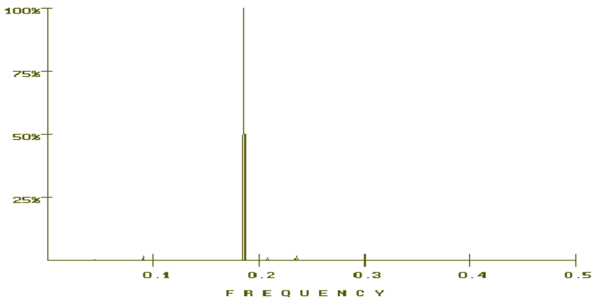

Figure 1 Cross-spectral function of gp160 env proteins from 19 different isolates revealing the characteristic frequency at f1=0.1855.

The 44 amino acid fragments of gp120, found to be crucial for binding to CD4, from 11 HIV isolates (lav, arv2, eli, ny5, wmj2, mal, z3, 3cg, cdc451, hxb2, hat) were also analysed. The RRM analysis showed that these regions share the same characteristic frequency f1 (as gp160env proteins), but they also have their own more prominent characteristic at frequency $f2=0.219\underline{+}0.022$ [1-3,12,13,15].

It is important to note that the frequency band f1=0.18 was also found to be significant in the spectra of p55gag and gp41env HIV proteins while frequency band f2=0.21 was found to be significant in the spectra of p17gag HIV proteins [1,2,4,12,13]. Thus both frequencies can be anticipated as a general characteristics of HIV proteins related to the viral interaction with CD4 as well as to its immune reactivity. To provide confirmatory evidence if one or both of those two characteristic frequencies characterise HIV immunoreactivity, *de novo* peptides having only f1 and/or f2 frequency characteristics in their RRM spectra have been previously designed using the RRM peptide design procedures [1-4,12-14]. These results have shown that synthetic peptides which are non-homologous to any HIV protein, but have only the same characteristic frequency as HIV env proteins, do generate the immune response which can recognise viral proteins [1,2,4,12,13]. The experimental confirmation that frequencies are important parameter for biomolecular recognition and in particular for antibody-antigen recognition has been shown by the significant cross-reactivity to the policlonal antibodies raised against peptides which share at least one characteristic frequency and phase [1,12,13].

Table 1:Characteristic RRM frequencies found in different HIV proteins.

Protien	freq.0	freq.1	freq. 2
Gp160env		0.18	
44mer120env		0.18	0.21
p55gag		0.18	
p17gag			0.21
Nef (HIV+SIV)	0.06		
T20			0.21
CD4 human			0.21
CD4 Ag (8)	0.035		
Gp120env	0.035	0.21	0.24
CD4 + gp120	0.035		

In addition, when all HIV and HIV target proteins were analysed, only 3 frequency bands have been identified pointing out that HIV virus is very simple in term of RRM frequencies (Table 1).

Interestingly when RRM was applied the frequency f2=0.21 was found to be characteristic of the interaction between CCR5 co-receptor and gp120env

IV RESULTS AND DISCUSSION

Three HIV inhibitory peptides, DP-107, peptide 637-666 and T20 peptide, have been identify so far as fragments of HIV envelope protein gp160. All of these HIV-1 inhibiting peptides selectively bind to the virus fusion domain, have same characteristic frequency as HIV-env proteins with phases opposite at this frequency [1]. As it was shown previously, HIV envelope proteins, gp160, from 19 HIV-1 isolates indeed have one common frequency f=0.185±0.001 (Fig.1). The same frequency was identified as common for gp41 and gp120 characterising their mutual interaction [1,2,4,12,13,15]. When inhibitory peptides were analysed all three of them have shown peak at frequency f=0.185±0.001 [1,15].

The other fragment of our interest is 44-mer gp120env fragment which was found to be crucial for binding to the CD4. All of these fragments are highlighted in the gp160 env sequence (Fig. 2). The braking point between gp120env and gp41env within the whole gp160env sequence is assigned with an arrow (↔).

We have shown previously [10,11,16] functional, active sites of proteins can be identified as high energy regions in the Continuous Wavelet Transform (CWT) of numerical presentation of the protein sequence. We applied here CWT to the whole envelope protein (gp160 - Fig. 3) and to its functional fragments (gp120 and gp41 figures 4,5 respectively). The aim is to see if it is possible to identify active CD4 binding site (44-mer fragment) as well as HIV inhibitory fragments (T20, DP-107 and 637-666 peptide) from CWT scalograms. Morlet wavelet function was used as it was shown earlier that it is best for use with protein sequences [16]. The characteristic frequency bands (f0=0.06, f1=0.18 and f2=0.21) are assigned in scalograms with arrows. It can be observed from the Figure 3 that there is a number of energetically significant domains in the whole gp160 HIV env sequence. However, as the sequence is relatively long (861 amino acids) there are too many details in the image and thus it is not easy to distinguish important details. More details can be observed from separate CWT scalograms for gp120 and gp41 respectively (Figures 4,5).

From gp120 scalogram (Fig. 4) it could be observed that the most significant area is from 430th to 480th amino acid, at the frequency f0, (circled in the figure). This fragment is in complete overlap with the 44mer fragments responsible for the binding to the CD4. Although these fragments by themselves do not have frequency f0 as the most significant CD4 antigen as well as the whole gp120 do have f0 as the most significant frequency (Table 1).

From gp41 scalogram (Fig. 5) four different significant regions could be observed. The first region is related to the frequencies f1 and f2 and is from about 540th to 570th amino acid. This region is overlapping inhibitory peptide DP107. The second region is related more to the frequency f0 is from about 590th to 630th amino acid. This region is between DP107 and 635-664 peptides with possible small overlaps with each of them. However, as this fragment is related more to the frequency f0 it would represent possible biding site rather than inhibitory peptide. The third significant area is related to the frequencies f1 and f2 and is from about 620th to 660th amino acid and is in complete overlap with both inhibitory peptides 637-666 and T20. Thus both significant regions that are related to the frequencies f1 and f2 are related to the inhibitory peptides as well. As it was found earlier [1,15] that inhibitory peptides are characterised by these frequencies the result is expected. There are no other peptidic candidates for inhibitory peptides according to the presented scalograms. However, there is a very strong and wide, fourth, region in the gp41 sequence which is from about 770th to 830th amino acid. This region is more related to the frequency f0 and thus could be a good candidate for the binding region. All significant regions in both gp120 and gp41 are circled in the related scallograms, figures 4 and 5.

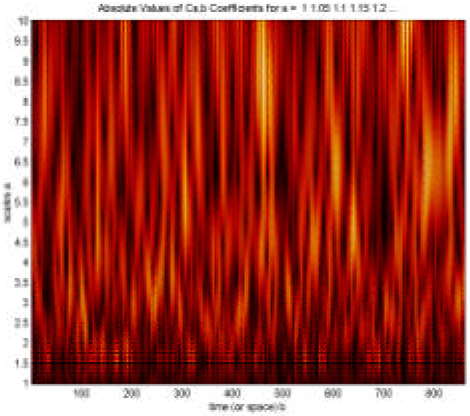

Figure 3: CWT of the whole gp160 HIVEnv strain lav (scale 10)

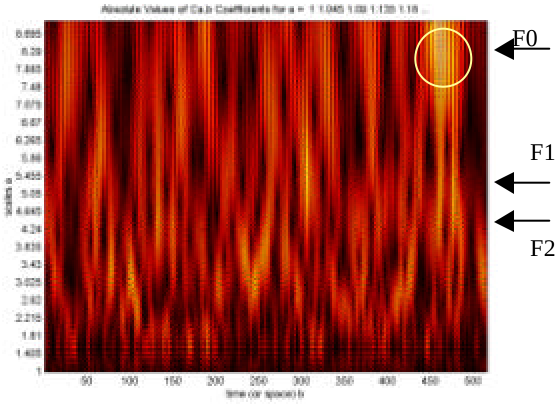

Figure 4: CWT of the gp120 HIVenv (scale 9)

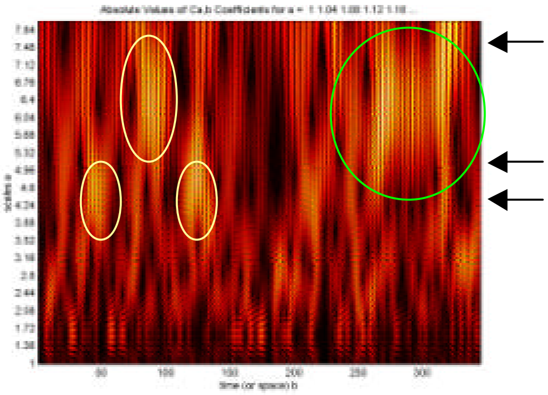

Figure 5: CWT of the gp41 HIVenv (scale 8)

V CONCLUSION

The RRM model has been applied here to analyse characteristics of HIV envelope proteins with the aim to identify functionally important fragments. Two separate groups of fragments have been identified:

- Three fragments which are related to frequency f0 (430-480, 590-630 and 770-830) . One of them is overlapped with the well known 44-mer fragments responsible for binding with CD4 antigen while the other two have still unknown function but they could be a good candidates for additional binding sites.
- Two fragments which are related to the frequencies f1/f2. Both of them are overlapped with experimentally established inhibitory peptides.

Thus it is possible to conclude that frequency f0 is more related to the binding with CD4 while frequencies f1 and f2 are more related to inhibition of this binding.

Having identified protein characteristics which are crucial for HIV interaction and inhibition is an important step towards understanding the HIV activity and designing the potent inhibitors.

ACKNOWLEDGEMENT

This work was supported by Monash Engineering Research Grant.

REFERENCES

1. Cosic I., (1997), The Resonant Recognition Model of Macromolecular Bioactivity: Theory and Applications, Birkhauser Verlag, Basel.
2. Cosic I., (1994), IEEE Trans. Biomedical Engineering, 41, 1101-1114.
3. Cosic I., (1995), Bio/Technology, 13, 236-238.
4. Cosic I., (1990), "Resonant Recognition Model of Protein-Protein and Protein-DNA Interaction", in Bioinstrumentation and Biosensors, ed Wise D., Marcel Dekker, inc, New York, 475-510.
5. Jiang S. et al., (1993), Nature, 365, 113.
6. Collier, N.C., Knox K., Schlesinger M.J., (1991), Virology, 183, 769-772
7. Kilby J.M. et al, (1998), Nature Med. 4:1302,
8. Veljkovic, I. Slavic, (1972), Physical Review Lett., 29: 105-108.
9. Akay M., (1998), Time Frequency and Wavelets in Biological Signal Processing, IEEE Press.
10. Fang Q., Cosic I., (1998), APESM, 21, No 4, 179-185.
11. De Trad C. H., Fang Q., **Cosic I.,** 2000, Biophysical Chemistry, 84/2, 149-157.
12. Krsmanovic V., Cosic I.,. Biquard J.M, Hearn M.T.W., (1998), Analogous Peptides of the Internal Image of a Viral Protein, Australian Patent AU-B-36361/93, acceptance number 682304
13. Krsmanovic V., Biquard J.M., Sikorska-Walker M., Cosic I., Desgranges C., Trabaud M.A., Whitfield J.F., Durkin J.P., Achour A., Hearn M.T., 1998, J.Peptide Res., 52(5), 410-4120
14. Cosic I., Drummond A.E., Underwood J.R., Hearn M.T.W., (1993), Molecular and Cellular Biochemistry, 130, 1-9.
15. Cosic I., Okada n., Okada H., (2001), IEEE EMBS Victorian Section, Melbourne, 120-123.
16. Cosic I, deTrad C., Fang Q., Akay M., 2000, IEEE-EMBS Asia-Pacific Conference, Hangzhou, China, 405-406

```
env LAV gp160(HIV) (CELL 45, 637, 1986
,Starcich,
B. R . et.al.)
MRVKEKYQHLWRWGWKWGTMLLGILMICSATEKLWVTVYYGVPVWKE
ATT

TLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWK
NDM

VEQMDEDIISLWDQSLKPCVKLTPLCVSLKCTDLGNASNTNSTNTNS
SSG

EMMMEKGEIKNCSFNISISIRGKVQKEYAFFYKLDIIPIDNDTTSYT
LTS

CNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTNV
STV

QCTHGIRPVVSTQLLLNGSLAEEEVVIRSANFTDNAKIIIVQLNQSV
ETN

CTRPNNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNATL
KQI

ASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
STW

FNSTWSTEGSNNTEGSDTITLPCRIKQFINMWQEVGKAMYAPPISGQ
IRC
```